

# AI를 활용한 건설 현장 스마트 관리 시스템

2025.07

팀 자동사냥 문운호(redhat1120@naver.com) 이지엽(olzlduq1@gmail.com) 이기표(kipyo95@naver.com)







## 목 차



- I 프로젝트 개요
- 개발 환경 및 적용 기술
- Ⅲ 데이터 수집 & 전처리
- 모델선정 및 평가
- ₩ 서비스 시연

# Part 01

프로젝트 개요

### Part 1 프로젝트 배경 및 목적 [1/3]



#### 건설업종 사망사고 증가…추락·붕괴, 화재 예방

제7차 현장점검의 날 계기…사고

2025.04.09 | 고용노동부 7

최근 건설업종에서 사망사고가 증. 생하고 있다.

이에 고용노동부와 한국산업안전<u>보</u> 고예방에 집중한다고 밝혔다.

특히 고용부는 전국 지방고용노동 험성이 높은 건설현장을 불시방문

2분기 건설현장서 63명 사망...작년보다 11명 증가

음 김윤태 기자 │ ② 승인 2023.0726 11:09 │ ഈ 댓글 0

올해 2분기 건설현장에서 사망한 노동자는 63명으로 지난해 같은 로 나타났다.

국토교통부는 지난 2분기 건설사고 사망자가 발생한 시공능력평기 책임이 있는 관련 하도급사, 공공발주 공사의 발주청, 인·허가기관( 다.

국토부가 건설공사 안전관리 종합정보망(CSI) 통계를 분석한 결과생한 사망자는 11개사 13명으로 3명 증가했다.

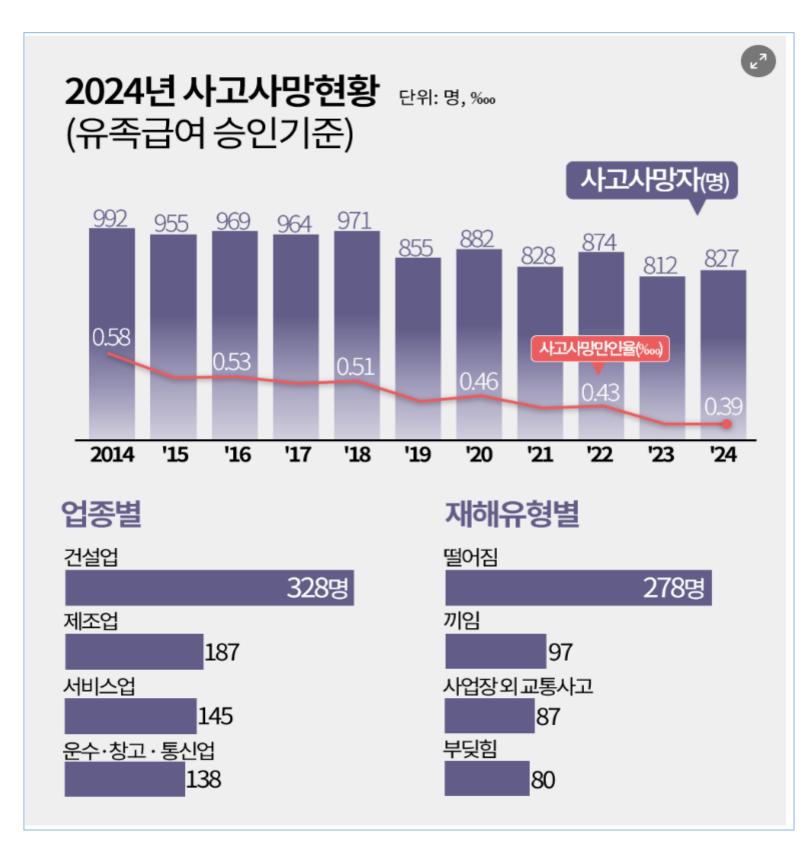
< 상위 100대 건설사(하도급사 포함)	시망사고 발생	<b>현황</b> (123,4~6월,	시골능력평가 순) >
------------------------	---------	----------------------	-------------

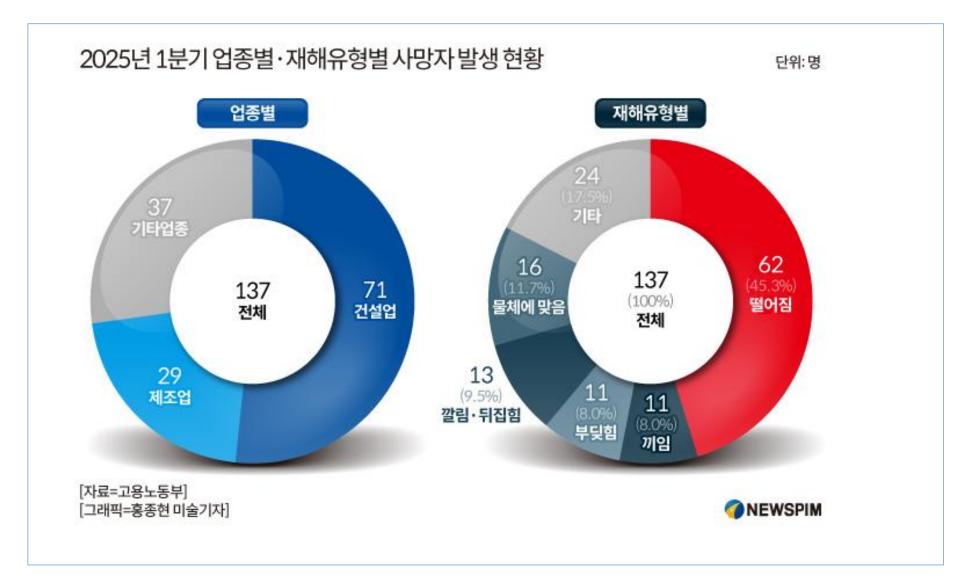
시공사		시공   - 능력   사고		사망자 사고일		
원도급	하도급	평가 종류 (*22년)				<u>공사</u> 명
현대	대천건축	7	떨어짐	1	4월 6일	현대 테라타워 <u>가산DK</u> 신축공사
<u>엔지니어링</u>	원영건업	'	떨어짐	1	4월 11일	<u>구리갈매</u> 지식산업센터 신축공사
롯데건설	케이이씨	8	떨어짐	1	5월 22일	청담삼익아파트 주택재건축정비사업
한화	한설그린	13	물체에 맞음	1	5월 10일	인천국제공항 <u>국제업무지역</u> (IBC)-III 개 발사업 1단계 <u>기반시설</u> 조성공사
진화	도양기업	13	물체에 맞음	1	5월 22일	고속국도 제29호선 세종~안성 간 건설공사(제2공구)
금호건설	안현건설	15	떨어짐	1	6월 3일	행정중심복합도시 6-3생활권 <u>H2BL</u> 주상 복합 신축공사
신세계건설	금계산업	34	끼임	1	5월 31일	스타필드 수원 신축공사
대광건영	유광토건	51	깔림	1	6월 15일	충북 오송 <u>BO8BL</u> 대광로제비앙 아파트 신축공사
시티건설	일광건설	57	깔림	1	5월 29일	평택~부여~ <u>의산(서부내륙)고속도로 민간</u> <u>투자사업</u> 제12공구
보미건설	청오중공업	71	떨어짐	1	5월 2일	<u>남양주물류센터</u> 신축공사
대우산업개 발	디더블유씨 앤디	75	떨어짐	1	4월 22일	서울 정류동 도시형생활주택 신축공사
동문건설	지승건설	76	떨어짐	1	4월 11일	고양 화정동 오피스텔 신축공사
미수건설	시온건영	94	떨어짐	1	5월 30일	주공생연아파트 재건축정비사업 신축공 사
	합.계			13		

[자료=국토교통부]

# Part 1 프로젝트 배경 및 목적 [2/3]







[출처: 고용노동부] https://www.newspim.com/news/view/20250521000258

[출처: 고용노동부] https://www.anjunj.com/news/articleView.html?idxno=40388

# Part 1 프로젝트 배경 및 목적 [2/3]



#### 기본적인 안전인식에 관한 문제





# Part 1 프로젝트 배경 및 목적 [3/3]



#### 현장점검 유무에 따른 사망 발생률 비교



[출처: 국토교통부] https://www.news1.kr/realestate/general/5593664

# Part 1 프로젝트 배경 및 목적 [3/3]



#### 현장점검 유무에 따른 사망 발생률 비교



[출처: 국토교통부] https://www.news1.kr/realestate/general/5593664



추진기간 (6월 25일 ~ 7월 18일)								
1주차	2주차							
데이터 수집								
전처리&가공								
NLP 데이터 전처리								
LLM 모델 설계								
	결과 분석							
	사고 대응 모델설계							
	딥러닝 모델 설계							
	결과 취합 및 발표 자료 작성							
	1주차 데이터 수집 전처리&가공 NLP 데이터 전처리							

### Part 1 요구사항 분석



요구사항 분석

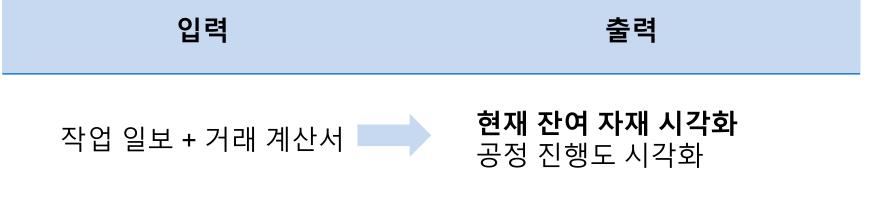
① 기능: 제반서류자동화

② 기능 :업무 진행률(변화량) 예측

③ 기능 : 업무진행정도 파악 자동화 잔여자재 시각화



출력 일보 + @ **주간 / 월간 진행률 시각화** 







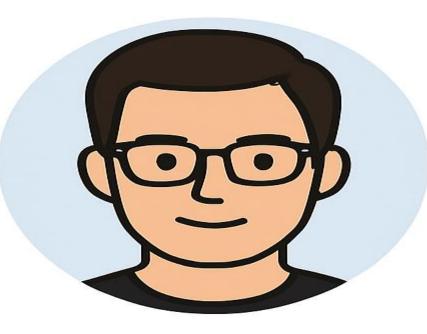
팀장 | 문운호

- 프로젝트 총괄 담당
- 모델전처리



팀원 | 이지엽

- 백엔드 기능구현
- 아키텍처
- 개발 아이디어 제공



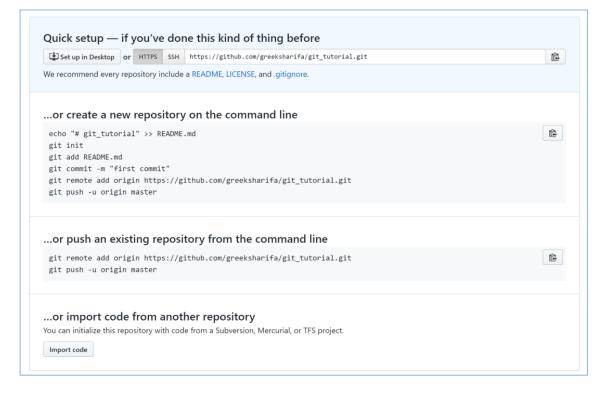
팀원 | 이기표

- 자료조사
- 프론트 담당
- 백엔드 보조

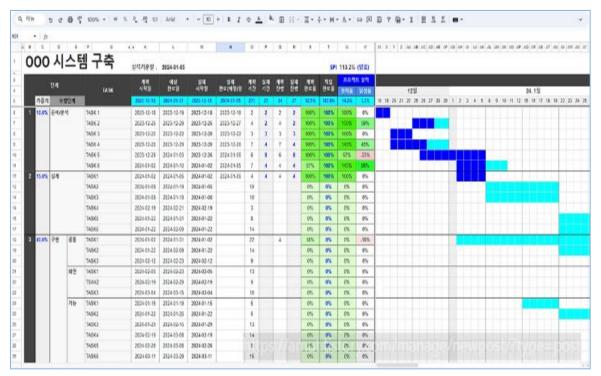
### Part 1 설계 및 협업



#### ■ 깃허브



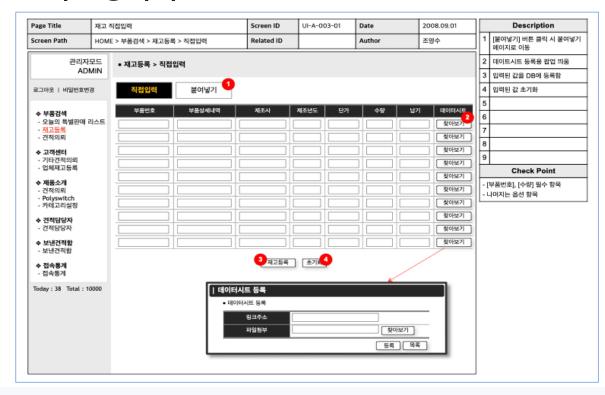
#### **■ WBS**



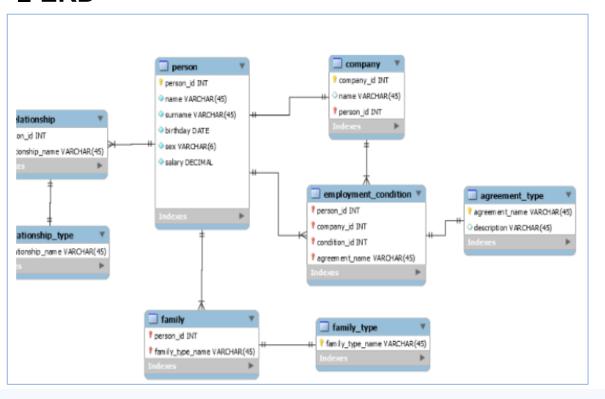
#### ■ 노션



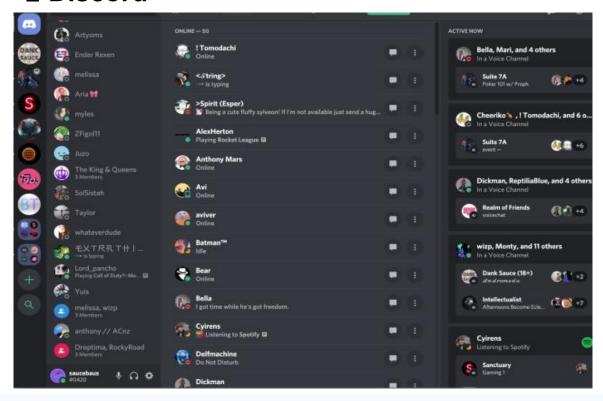
#### ■ 화면정의서



#### **■ ERD**



#### **■** Discord

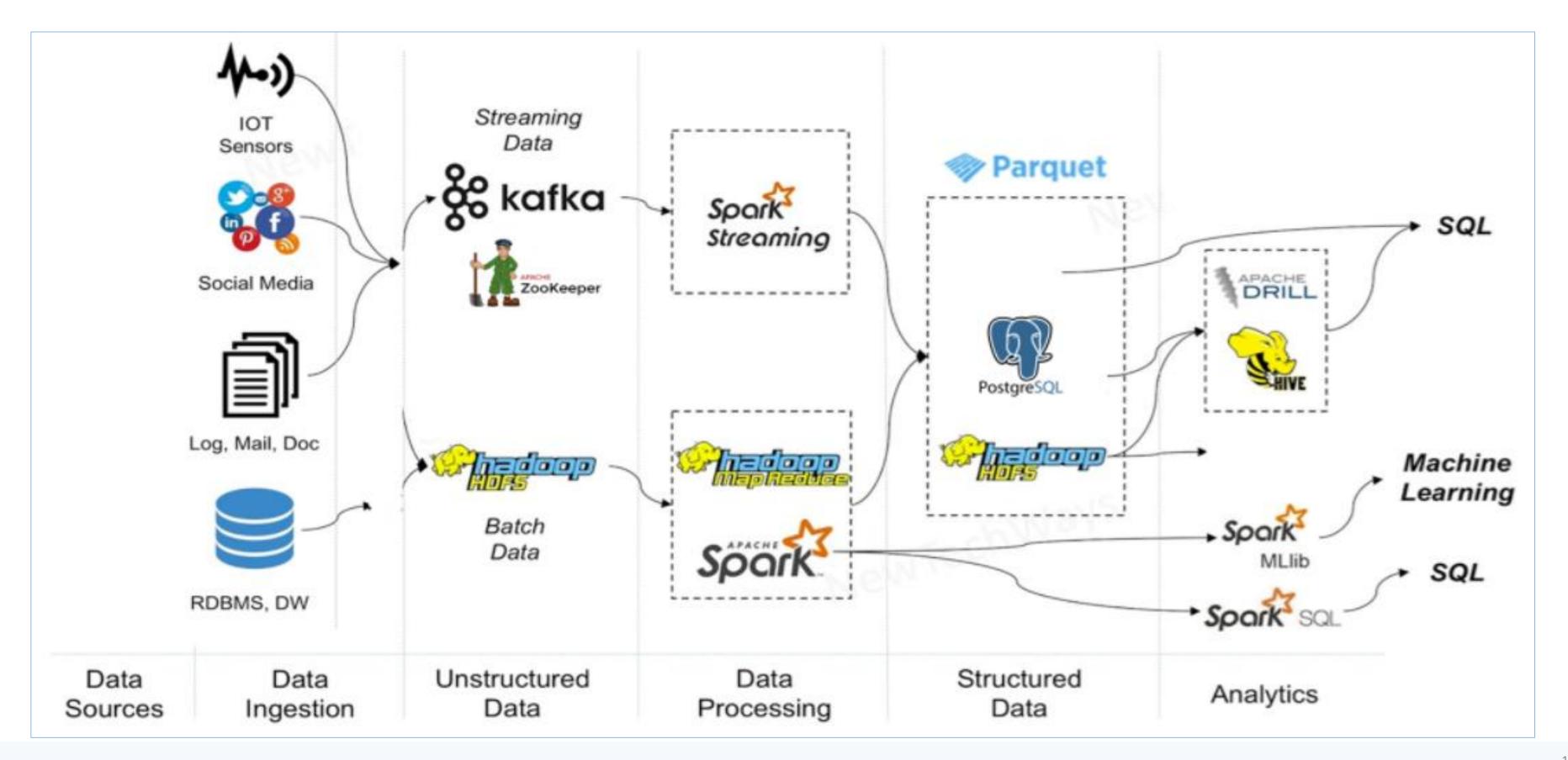


# Part 02

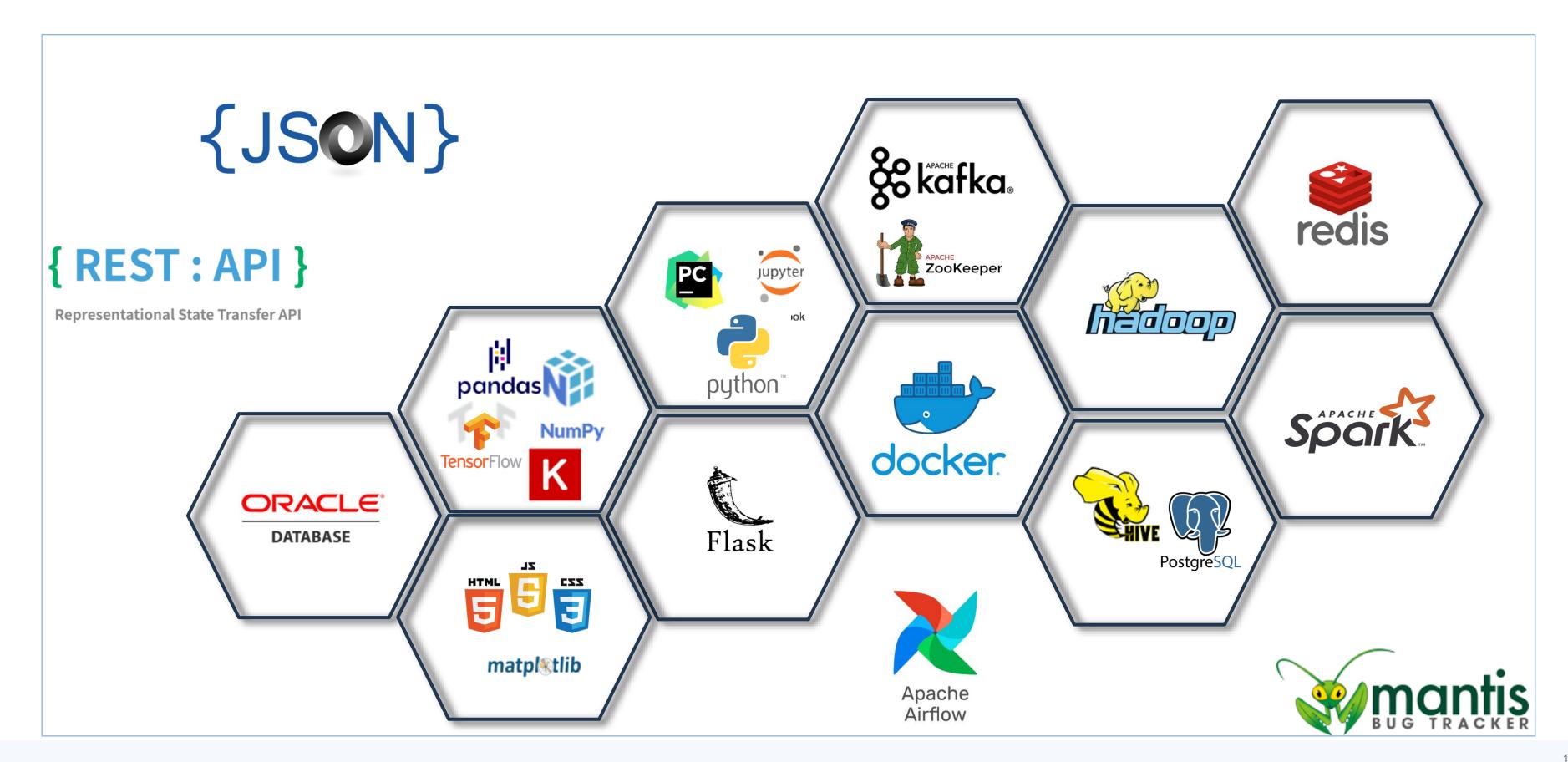
개발환경및적용기술

## Part 2 개발환경









# Part 03

데이터수집및전처리

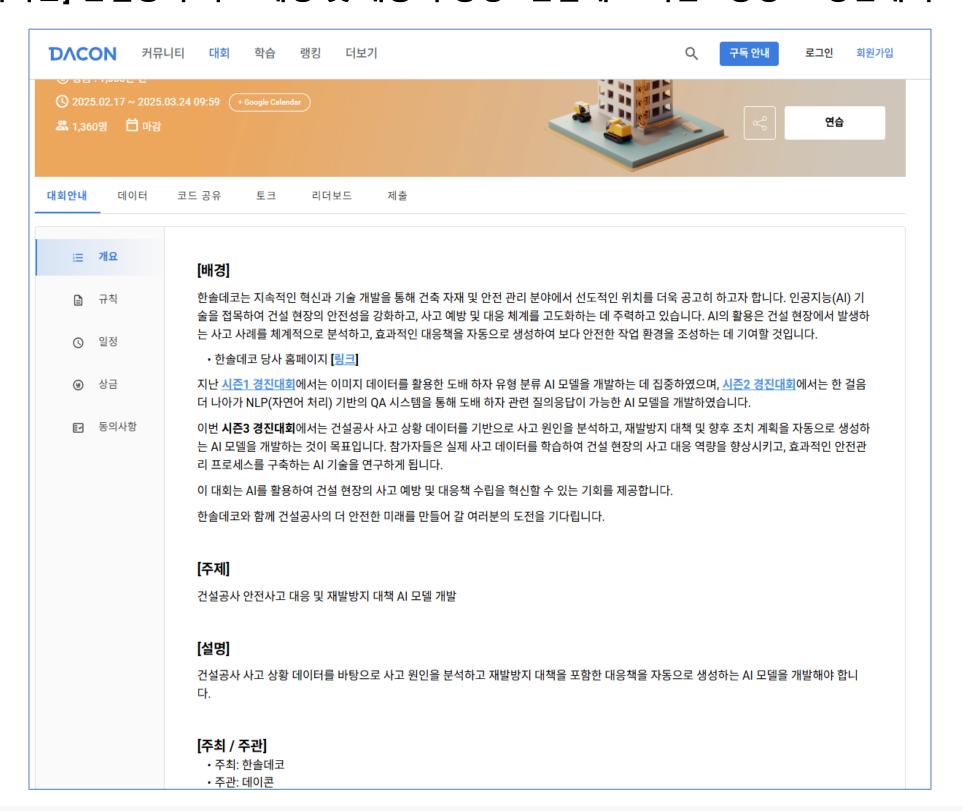
# Part 3 데이터셋 [1/2]



#### [국토부] 건설현장 주요사고 예방 공모전

#### 국토교통부 💠 국토안전관리원 공모전 안내 공지사항 일시: 2025.07.16-2025. 08.26 수상발표 : 2025.09. 16. (화) 공모대상 : 대한민국 국민 누구나 2025. 7. 16. ~ 8. 26 조요사고 예방 건설현장에서 빈번하게 발생하는 추락 및 깔림사고를 예방하기 위하여 국민 공모를 통해 선정된 'TBM 교육카드(20종)', '사고예방 슬로건'의 확산 챌린지를 시작합니다! 총 상금 600만원의 주인공이 되어주세요! 대국민 참여 부문 사고예방 슬로건 외치기 챌린지 TBM 교육카드 영상 콘텐츠 공모전 바로가기 공모전 종료 까지 36 08 56 39 대국민 참여 부문 사고예방 안전모 밀착 건설 안전에 관심있는 국민 누구나 ('개인'으로 참가) 슬로건 외치기 \* 영상 내 참여인원 1인 이상, 개인당 최대 1번 제출 가능합니다 안전대 장착 챌린지 무재해 정착 건설현장의 안전을 기원하는 나(우리)의 외침을 자유롭게 촬영 후 제출해주세요 \* 슬로건이 적힌 인쇄물 혹은 손글씨 피켓을 들고 이목구비가 명확하게 나와야 인정 '안전모 밀착, 안전대 장착, 무재해 정착!' 시상내역 총 85건, 총 상금 300만원 다함께 외쳐, 건설현장의 안전을 만들어 나가요 선착순상 (75인) 2만원 상당 상품권 지급 SNS 특별상 (10인) 15만원 (SNS 미션을 달성한 참여자 중 추첨) \* SNS 미션 1. SNS 해당 영상 '좋아요' 20개 또는 '조회수' 200회 이상 달성 2. 영상 제출 시 SNS에 업로드한 링크를 함께 제출하면 참여 완료! (전체공개 필수)

#### [데이콘] 건설공사 사고 예방 및 대응책 생성 : 한솔데코 시즌3 생성 AI 경진대회



# Part 3 데이터셋 [2/2]



- 분석 대상 기간 : 2020.01 ~ 2023.05

- 건설 사고 데이터 : 23427(rows) \* 18(cols)

- 주요 컬럼 : 발생일시, 사고원인, 공종, 작업프로세스, 인적사고, 물적사고, 장소, 사고객체, 부위

ID	발생일시	사고인 지 시간	날 씨 기온	습 도	공사종류	연면적	층 정보	인적사고	물적 사고	공종	사고객체	작업프 로세스	장소	부위	사고원인	재발방지대책 및 향후조치계획
0 TRAIN_00000	2023-12-31 오후 12:44	정규작 업 -	망 음 1°0	30%	건축 / 건축물 / 근린생활시설	4,892.77m²	지상 14층, 지하 3층	떨어짐(5미터 이 상 ~ 10미터 미 만)	없음	건축 > 철근 콘크리트공사	건설자재 > 철근	설치작 업	근린생활시 설 / 내부	철근 / 고 소	고소작업 중 추락 위험이 있음에도 불구하 고, 안전난간대, 안전고리 착용 등 안전장 치	고소작업 시 추락 위험이 있는 부위에 안전 장비 설치.
1 TRAIN_00001	2023-12-30 오후 03:35	정규작 업 -	맑 음 10°0	90%	토목 / 터널 / 철도터널	-	-	끼임	없음	토목 > 터널 공사	건설자재 > 볼트	운반작 업	철도터널 / 내부	볼트 / 바 닥	부주의	재발 방지 대책 마련과 안전교육 실시.
2 TRAIN_00002	2023-12-30 오후 02:30	정규작 업 -	맑 음 14°0	70%	건축 / 건축물 / 업무시설	1,994.62m²	지상 5층, 지하 0층	넘어짐(미끄러짐)	없음	건축 > 철근 콘크리트공사	기타 > 기 타	형틀 및 목공	업무시설 / 내부	기타 / 바 닥	3층 슬라브 작업시 이동중 미끄러짐	현장자재 정리와 안전관리 철저를 통한 재 발 방지 대책 및 공문 발송을 통한 향후 조
3 TRAIN_00003	2023-12-30 오후 12:00	휴일근 무 -	호 림 12°0	55%	토목 / 하천 / 기타	-	-	기타	전도	토목 > 하천 공사	부재 > 교 각 기초	조립작 업	기타 / 동산 교 신축구간	교각 기초 / 바닥	교각 기초철근 조립 중 강한 바람에 의해 기 둥측 주철근이 균형을 잃고 전도되어, 하	위험성 평가 및 교육을 통해 작업장 내 위험 요인과 안전수칙을 근로자에게 전파하고,
4 TRAIN_00004	2023-12-30 오전 10:00	정규작 업 -	맑 음	10%	건축 / 건축물 / 공동주택	59,388.93 m²	지상 27층, 지하 3층	넘어짐(미끄러짐)	없음	건축 > 해체 및 철거공사	기타 > 건 설폐기물	해체작 업	공동주택 / 내부	건설폐기 물 / 바닥	근로자의 부주의	자재 정리 작업 시 세부 작업 방법에 대한 교육 실시와 작업 구간 이동 경로 점검

#### 데이터 전처리 [1/3] Part 3



고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음



▼ 1. 토크나이저 (형태소 분석기)

['고소', '작업', '중', '추락', '위험', '이', '있음', '에도', '불구', '하고', ',', '안전', '난간대', ',', '안전', '고리', '착용', '등', '안전', '장치', '가', '미흡', '하였음']



2. Stopword + 특정 지역명 제거 단어가 아닌 토큰, 큰 의미가 없는 토큰 제거

['고소', '작업', '추락', '위험', '있음', '안전', '난간대', '안전', '고리', '착용', '안전', '장치', '미흡']



3. 빈도 처리 사전 빈도수 30%이하 제거

['고소', '작업', '추락', '위험', '안전', '장치', '미흡']



4.

4. Voca

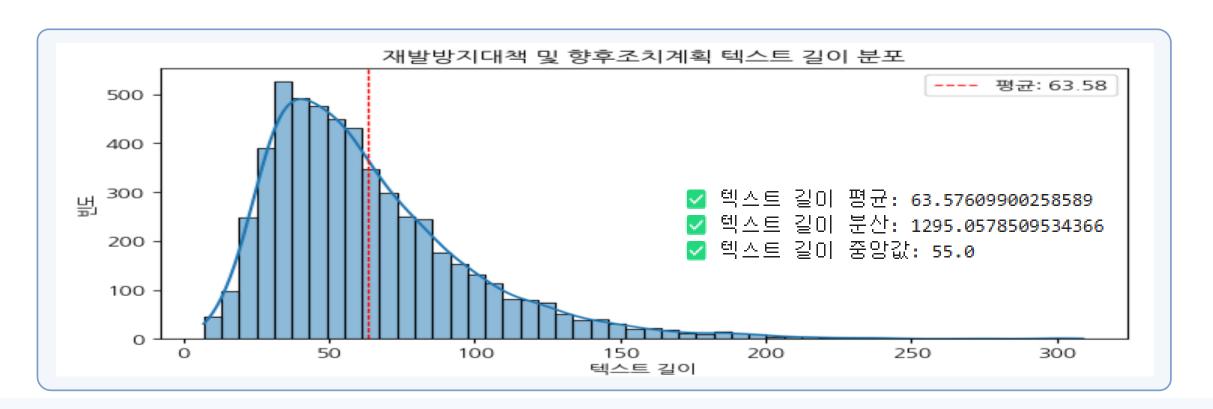
['고소', '작업', '추락', '위험', '안전', '장치', '미흡']

#### **Word Dictionary (Voca)**

1	가구	130	안전
2	가방		
7	고소	214	작업
		215	장치
128	미흡		
129	위험	613	추락

#### > 5. Word Embedding

[7, 214, 613, 129, 130, 215, 128]





#### 대상 컬럼

- 사고원인
- 공종
- 작업프로세스
- 인적사고
- 물적사고
- 장소
- 사고객체
- 부위

#### 데이터 벡터화

```
{'사고원인': array([[ 0.02095127, 0.06670206, 0.04533919, ..., -0.06525377,
        -0.06577135, -0.02643926],
        [ 0.0097357 , 0.05543886, -0.04375602, ..., -0.07560129,
       -0.02000579, 0.03043894]], dtype=float32),
'공종': array([[-0.01535195, 0.01893546, 0.01329352, ..., -0.05164649,
        -0.05076602, 0.0289134],
       [-0.01446563, 0.05385562, -0.00414923, ..., -0.05257048,
        -0.04301524, 0.0121932 ]], dtype=float32),
 '작업프로세스': array([[-0.00809287, -0.02479578, 0.02874802, ..., 0.02538222,
         -0.06783103, 0.05124296],
        -0.05816261, 0.04935975]], dtype=float32),
 '인적사고': array([[ 0.02596208, 0.00211758, -0.04276543, ..., -0.07385457,
        -0.01250095, -0.0237977 ],
        -0.0807321 , -0.03443997]], dtype=float32),
'물적사고': array([[ 0.01502613, 0.03261068, 0.01440841, ..., 0.05848887,
        -0.08067237, 0.04637586],
        -0.0807321 , -0.03443997]], dtype=float32),
 '사고객체': array([[-0.03149208, 0.0246365, 0.00131126, ..., -0.03103822,
         -0.0685368 , 0.01594386],
         -0.11339174, 0.06666487]], dtype=float32),
  '장소': array([[ 0.0124729 , 0.02483705, 0.02041617, ..., -0.06431791,
         -0.12960592, 0.03677965],
         -0.11339174, 0.06666487]], dtype=float32),
 '부위': array([[-0.00694059, 0.03287686, -0.01856954, ..., 0.01015181,
         -0.05988195, 0.00922823],
        -0.10686478, 0.03687347]], dtype=float32)}
```

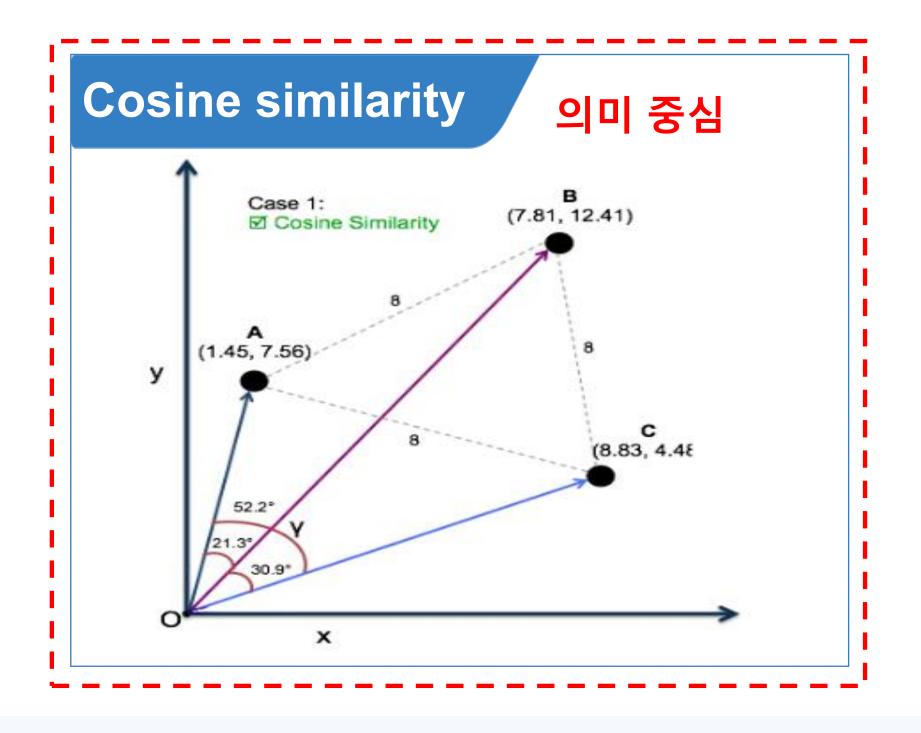
# Part 04

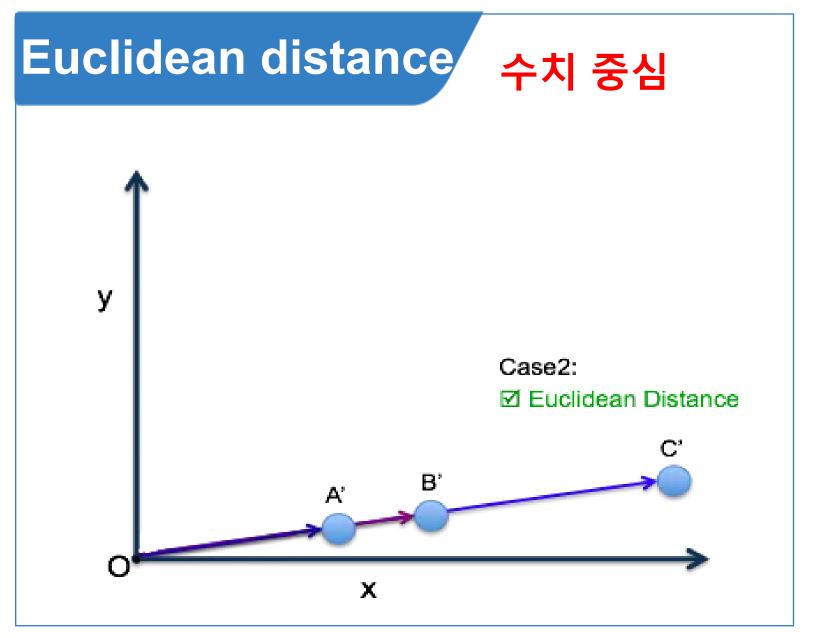
모델선정및평가

### Part 4 유사도 — 거리기반



• 각도(코사인) 또는 길이(유클리디안) 기준으로 벡터를 얼마나 가깝다고 볼 것인가?





### Part 4 가중치 선정 [1/2]



- Information Gain / Mutual Information (통계 기반 가중치 산정)
  - Mutual Information 계산
  - 값이 클수록 해당 feature가 타겟 예측에 더 유용
- 분류/예측 정확도 기반 튜닝 (지도학습 기반) → X 우리 데이터는 라벨이 없는 비지도학습

  - 분류/예측 정확도 기반 튜닝 (지도학습 기반)
    가중치를 다르게 조합하여 최종 유사도 행렬 생성

  - 이 최종 유사도를 기반으로 예측(분류 혹은 유사도 판단)을 수행
    예측 결과와 실제 라벨(ground truth)을 비교해 성능 평가 (예: 정확도, F1-score)

- 클러스터(군집) 품질 기반 가중치 산정
  - 각 feature별 유사도 행렬을 이용해 별도 클러스터링 수행
  - 군집의 응집도(cohesion), 분리도(separation), 실루엣 점수 등을 구함
  - 품질이 좋은 feature에 더 높은 가중치 부여



#### Part 4 Information Gain / Mutual Information (통계 기반 가중치 산정)

- 기준 사고원인: 고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음.
- 기준 공종: 건축 > 철근콘크리트공사
- 기준 작업프로세스: 설치작업
- 기준 장소: 근린생활시설 / 내부
- 기준 부위: 철근 / 고소
- ➡ 대응 대책: 고소작업 시 추락 위험이 있는 부위에 안전장비 설치.
- 1번째 유사한 사고 (유사도: 0.8733)
- 사고원인: 데크플레이트 설치작업 중 안전고리 등 개인 안전장비를 착용하고 작업하였으나, 데크 중앙부에서 작업을 다.
  - ◆ 공종: 건축 > 철골공사
  - ◆ 작업프로세스: 설치작업
  - ◆ 장소: 자동차 관련시설 / 내부
  - 부위: 철골부재 / 상부(위)
- ➡ 대응 대책: 작업 전 개별 안전장비 착용 여부 확인과 근로자 안전 교육 실시를 통한 사고 재발 방지 대책 마련.
- 2번째 유사한 사고 (유사도: 0.8681)
- 사고원인: 지하1층 가시설 해체작업중 안전모, 안전벨트, 안전화 등을 착용하였으나, 안전벨트 체결은 하지 못한 상태
- ◆ 공종: 토목 > 가설공사
- 작업프로세스: 해체작업
- ◆ 장소: 근린생활시설 / 내부
- 부위: 따장 / 고소
- ➡ 대응 대책: 안전벨트 체결을 위한 안전 시설물 설치와 선안전, 후시공 원칙 준수를 통한 근로자 안전 관리.
- ◆ 3번째 유사한 사고 (유사도: 0.8610)
- 사고원인: 고소작업 시. 개인보호구 착용을 철저히 하고 상부에 안전대 부착시설에 안전고리 체결 후 작업을 실시하며
- ◆ 공종: 토목 > 프리캐스트 콘크리트공사
- ◆ 작업프로세스: 해체작업
- ◆ 장소: 도로 / 외벽
- ◆ 부위: 철근 / 고소
- ➡ 대응 대책: 불안전한 행동 금지와 해당 공종 작업자 특별 교육 실시, 관리감독자의 현장 상주 및 작업 감독, 고소작업
- 4번째 유사한 사고 (유사도: 0.8605)
- 🔵 사고원인: 안전시설물 미설치(외부 시스템비계 설치 작업 시 자재를 들고 미동 중 무릎이 벽체 철근에 찔림)
- 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 공동주택 / 외부
- ◆ 부위: 철근 / 앞
- ➡ 대용 대책: 근로자 작업 및 이동을 위한 통행로 조성과 통행로 인근 철근에 철근캡 설치, 안전시설 설치에 대한 철저한
- 5번째 유사한 사고 (유사도: 0.8587)
- 사고원인: 안전모,안전밸트 착용 후 안전고리를 생명로프에 걸고 체육관 2층 철골기둥 작업 중 발이 미끄러지며 안전
- ◆ 공종: 건축 > 철골공사
- 작업프로세스: 설치작업
- 장소: 운동시설 / 내부
- ◆ 부위: 건물 / 바닥
- ➡ 대응 대책: 지속적인 안전교육과 수시 위험성 평가를 통한 위험 요소 제거로 사고를 미연에 방지하는 대책.

고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음.

기준 사고	안전장치 미흡	추락	유사도
유사1	임시로 안전고리를 해제 후 작업	낙상하는 사고가 발생	0.8733
유사2	안전벨트 체결은 하지 못한 상태	5m아래로 추락한 사 고	0.8681
유사3	안전고리 체결 후 작업을 실 시하여야 하나 근로자가 이 에 대해 부주의	발생한 사건	0.8610
유사4	자재를 들고 이동 중	무릎이 벽체 철근에 찔 림	0.8605
유사5	발이 미끄러지며 안전고리 가 빠져서	발생한 사고	0.8587

# Part 4 클러스터(군집) 품질 기반 가중치 산정



- 기준 사고원인: 고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음
- 기준 공종: 건축 > 철근콘크리트공사
- 기준 작업프로세스: 설치작업
- 기준 장소: 근린생활시설 / 내부
- 기준 부위: 철근 / 고소
- ➡ 대응 대책: 고소작업 시 추락 위험이 있는 부위에 안전장비 설치.
- 1번째 유사한 사고 (유사도: 0.9372)
- 사고원인: 1층 PC기둥 위에 PC거더 설치 작업 중 PC거더 위치 미세 조정을 위해 PC거더에 체결된 4개의 샤클 중 20고소작업대에 타고 있던 PC공이 약 9미터 아래 콘크리트 바닥으로 추락
- ◆ 공종: 건축 > 철근콘크리트공사
- ◆ 작업프로세스: 설치작업
- ◆ 장소: 창고시설 / 내부
- 부위: 거더 / 고소
- ➡ 대응 대책: 고용노동부의 부분공사중지(PC공사)명령에 따른 사고원인 보고서와 재발방지 계획의 발표 후 후속조치
- 2번째 유사한 사고 (유사도: 0.9352)
- 사고원인: 안전대고리를 걸고 작업을 했어야 하나, 보 거푸집 제작이 잘못되어 거푸집을 수정하기 위하여 폼 판넬의
- 공종: 건축 > 철근콘크리트공사
- ◆ 작업프로세스: 설치작업
- ◆ 장소: 공장 / 내부
- ◆ 부위: 옹벽 / 고소
- ➡ 대응 대책: 안전대 착용, 안전로프 설치, 안전고리와 동바리 수평재 연결, 유해 위험요인 인지 교육 및 안전 수칙 |
- 3번째 유사한 사고 (유사도: 0.9345)
- 사고원인: 데크플레이트 설치작업 중 안전고리 등 개인 안전장비를 착용하고 작업하였으나, 데크 중앙부에서 작업· 다.
- ◆ 공종: 건축 > 철골공사
- ◆ 작업프로세스: 설치작업
- 장소: 자동차 관련시설 / 내부
- 부위: 철골부재 / 상부(위)
- ➡ 대응 대책: 작업 전 개별 안전장비 착용 여부 확인과 근로자 안전 교육 실시를 통한 사고 재발 방지 대책 마련.
- 4번째 유사한 사고 (유사도: 0.9331)
- 사고원인: 철골 상부 작업시 추락방지용 안전고리를 체결했지만 생명줄 길이가 길어 추락 방지 효과 미발현
- 광종: 건축 > 금속공사
- 작업프로세스: 설치작업
- ◆ 장소: 업무시설 / 내부
- 부위: 자재 / 고소
- ➡ 대응 대책: 현장 내 고소작업 시작 전 관리자에 의한 생명줄 길이 점검과 고소작업자 특별안전교육 실시.
- ◆ 5번째 유사한 사고 (유사도: 0.9323)
- 사고원인: 재해자와 크레인 기사와의 무선신호 주파수가 맞지 않아 의사소통의 문제로 보거푸집이 결박에서 양중되
- ◆ 공종: 토목 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 근린생활시설 / 내부
- ◆ 부위: 철근 / 바닥
- ➡ 대응 대책: 작업전 무선신호 교신상태 재확인, 거푸집 작업지휘자 배치, 작업전 안전교육 실시.

고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음.

기준 사고	안전장치 미흡	추락	유사도
유사1	PC거더 전복	콘크리트 바닥으로 추 락	0.9372
유사2	폼 판넬의 못을 제거하던 중	떨어지는 재해 발생	0.9352
유사3	임시로 안전고리를 해제 후 작업	낙상하는 사고가 발생	0.9345
유사4	생명줄 길이가 길어	추락 방지 효과 미발현	0.9331
유사5	보거푸집이 결박에서 양중 되어	추락함	0.9323



#### • 클러스터(군집) 품질 기반 가중치 산정

- 대상 feature 별 유사도 행렬 계산
- 유사도를 거리로 변환 (거리 = 1 유사도)
- MDS를 통해 유사도 기반 거리 행렬을 2차원 임베딩
- KMeans 클러스터링 수행
- 실루엣 점수로 군집 품질 평가
- 클러스터 개수(K) 튜닝 : 최적의 K=5
- 가중치는 총합 1로 정규화하여 계산에 반영

```
# 가중치 설정
weights = {
 '사고원인': 0.0745, '공종': 0.1454, '작업프로세스': 0.1247, '인적사고': 0.1353,
 '물적사고': 0.2127, '장소': 0.1182, '사고객체': 0.1078, '부위': 0.0814
}
```

### Part 4 유사도 — TF-IDF



- 거리기반 유사도 vs. 단어빈도를 사용한 유사도 중 어느 것이 성능이 좋을까?
  - Cosine similarity → TF-IDF 로 변경
- 가중치는 기존의 "클러스터(군집) 품질 기반 가중치 산정" 그대로 사용

```
# 가중치 설정
weights = {
    '사고원인': 0.0745, '공종': 0.1454, '작업프로세스': 0.1247, '인적사고': 0.1353,
    '물적사고': 0.2127, '장소': 0.1182, '사고객체': 0.1078, '부위': 0.0814
}
```

## Part 4 유사도 — TF-IDF



- 기준 사고원인: 고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음 ➡ 대응 대책: 고소작업 시 추락 위험이 있는 부위에 안전장비 설치.
- 1번째 유사한 사고 (유사도: 0.8696)
- 사고원인: 도로에서 볼때 우측에서 4번째 기둥의 철근을 조립하는 과정에서 목격자 신\*\*은 수직철근을 바닥에 대고속하던 중 발이 미끄러지면서 시스템비계와 건물사이 틈으로 추락
  - 공종: 건축 > 철근콘크리트공사
- ◆ 작업프로세스: 설치작업
- 장소: 근린생활시설 / 외부
- 부위: 철근 / 상부(위)
- ➡ 대용 대책: 작업자 한전주의 교육 철저와 사고 작업자 치료완치 후 작업 투입 조치로 동일 사고 발생 방지.
- ◆ 2번째 유사한 사고 (유사도: 0.8103)
- 사고원인: 작업중 올라가려다 미끄러짐
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 근린생활시설 / 내부
- ◆ 부위: 철근 / 바닥
- ➡ 대응 대책: 높은 곳 작업 시 안전한 작업발판 설치와 안전고리 사용, 성급한 작업 지양 및 안전교육 실시를 통한 안 사항 재발 방지를 위한 수시 안전교육 실시.
- 3번째 유사한 사고 (유사도: 0.7594)
- 사고원인: 1. 지하 4층 B구간 E/V실 피트 내부에서 철근배근을 위해 철근을 바닥에서 들어올리던 중 하부에 설치 된
- 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 근린생활시설 / 내부
- 부위: 철근 / 바닥
- ➡ 대응 대책: 근로자의 부주의 행동 방지를 위한 관리 및 교육, 공종사고 예방을 위한 전파 교육과 작업장 방호조치 :
- 4번째 유사한 사고 (유사도: 0.7578)
- 사고원인: 지하 2층 기계실 기둥 철근 조립 공정에서 시스템 동바리 상부에 설치된 작업발판 상부에 놓여진 자재(유
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 공동주택 / 내부
- 부위: 철근 / 지하
- ➡ 대응 대책: 기둥철근조립용 작업대 설치, 차락 및 낙하방지망 설치, 해당 공종 작업자에 대한 안전교육 실시와 위험
- 5번째 유사한 사고 (유사도: 0.7558)
- 사고원인: 철근 작업자가 PIT층에서 철근 절곡기를 이용한 절단작업 중 절단된 철근(약 3cm)이 인중으로 튀면서 살
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- 장소: 근린생활시설 / 내부
- 부위: 철근 / 상부(위)
- ➡ 대응 대책: 철근 짧게 절단 시 튈 위험이 있으므로 보안면 착용 후 작업을 실시하고, 귀가조치 후 명일 정상 출근 예

고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음.

기준 사고	안전장치 미흡	추락	
유사1	발이 미끄러지면서	건물사이 틈으로 추락	0.8696
유사2	작업중 올라가려다	미끄러짐	0.8103
	취고 6 메디에 디 트 에 6 리	프레이트 미니리에 파	
유사3	철근을 바닥에서 들어올리 던 중	플레이트 모서리에 팔 뚝을 긁히는 사고 발생	0.7594
유사3  유사4			0.7594

# Part 4 중간정리



01

#### 분석 데이터셋

- 공사현장 사고 데이터
- 자재관리 이력 데이터
- 근로자 출근 이력

02

#### 전처리/가공

- 사전만들기
- 임베딩

03-1

#### 유사도 계산

- Cosine similarity
- TF-IDF

**Dot Product** 

03-2

가중치 계산

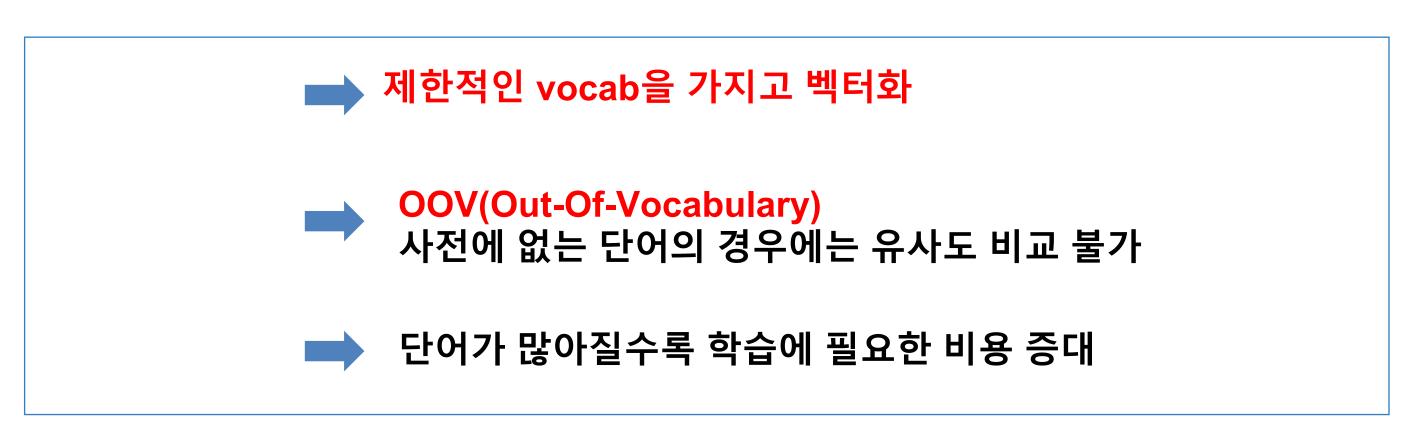
- MI 통계 기반 가중치
- 군집 기반 가중치

최종 유사도 행렬

### Part 4 문제점(개인학습한 사전의 문제점)



• Train 데이터를 임베딩해 vectorize 할 경우 문제점



#### WordPiece 임베딩 필요

"서브워드 단위의 토큰"을 사전에서 찾아 쪼개는 방식



사전 학습된 큰 임베딩 사전 필요

# Part 4 중간정리



01

분석 데이터셋

- 공사현장 사고 데이터
- 자재관리 이력 데이터
- 근로자 출근 이력

전처리/가공

- 사전만들기
- 임베딩

03-1

유사도 계산

- Cosine similarity
- TF-IDF

**Dot Product** 

03-2

가중치 계산

- MI 통계 기반 가중치
- 군집 기반 가중치

02

사전학습된 KoBERT 임베딩 도입

기존 train 데이터를 사용한 사전 → 위키피디아를 사용한 사전

최종

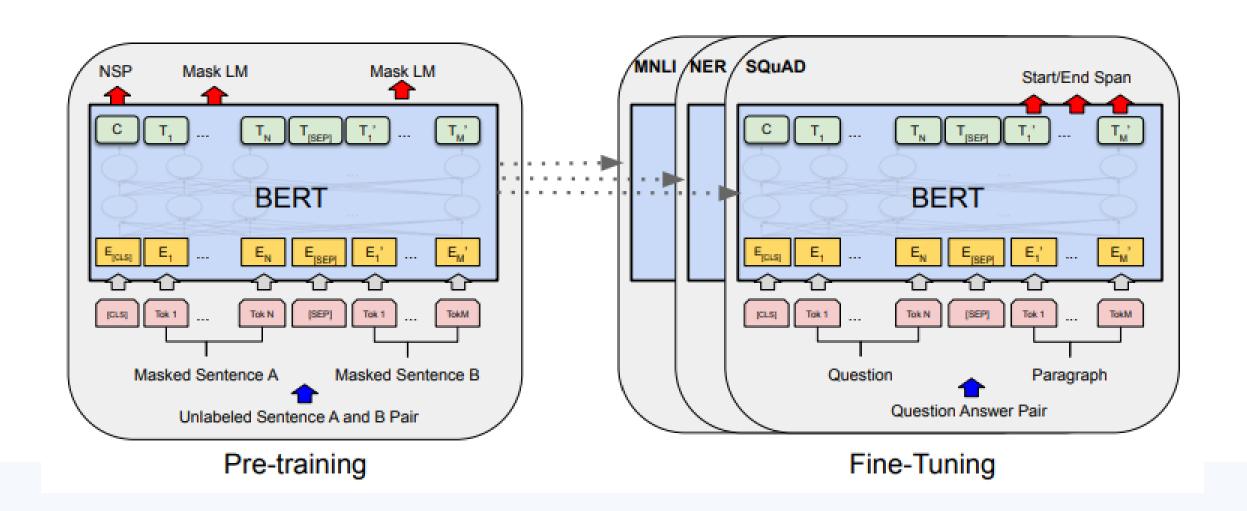
유사도

행렬

### Part 4 KoBERT 임베딩 기반 유사도 계산



- BERT (Pre-training of Deep Bidirectional Trnasformers for Language Understanding)
  - Transformer의 Encoder 모델을 쌓아 만든 모델
  - 사전 학습된 모델을 fine-tuning 해서 사용
- Kobert
  - \_ 기존 BERT의 한국어 성능 한계를 극복하기 위해 개발
  - \_ 위키피디아나 뉴스 등에서 수집한 수백만 개의 한국어 문장으로 이루어진 대규모 말뭉치 사전 학습



### Part 4 KoBERT 임베딩 기반 유사도 계산



- 기준 사고원인: 고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음
   ➡ 대응 대책: 고소작업 시 추락 위험이 있는 부위에 안전장비 설치.
- 1번째 유사한 사고 (유사도: 0.9597)
- 사고원인: 조립이 진행중인 강관비계 작업발판 위에서 안전난간대 설치완료 전까지의 추락방호를 위한 작업 중 조계
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 근린생활시설 / 내부
- 부위: 기타 / 바닥
- ➡ 대응 대책: 사고사례 전파 및 관련 특별안전교육 시행, 일일 작업전 안전점검 및 안전교육 철저 시행으로 작업자 9
- 2번째 유사한 사고 (유사도: 0.9550)
- 사고원인: 거푸집 해체공이 발판 해체 후, 완전히 고정을 못함
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 종교시설 / 외벽
- 부위: 작업발판 / 고소
- ➡ 대응 대책: 안전용품 철저와 작업 전후 체결 확인을 통한 재발 방지 대책 및 발판 해체 후 고정하지 않을 시 현장
- 3번째 유사한 사고 (유사도: 0.9544)
- 사고원인: 1층 바닥보 데크플레이트 판개작업 중 현 위치에 맞는 자재가 없어 다른 자재를 가져오던 중 강풍에 못여
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 교육연구시설 / 외부
- ◆ 부위: 기타 / 고소
- ➡ 대응 대책: 고소작업 시 안전벨트 결속을 위한 특별안전교육 실시와 안전대 부착설비 설치 및 안전대 착용 상태에서
- 4번째 유사한 사고 (유사도: 0.9527)
- 사고원인: 안전대고리를 걸고 작업을 했어야 하나, 보 거푸집 제작이 잘못되어 거푸집을 수정하기 위하여 폼 판넬의
- 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 설치작업
- ◆ 장소: 공장 / 내부
- 부위: 옹벽 / 고소
- ➡ 대응 대책: 안전대 착용, 안전로프 설치, 안전고리와 동바리 수평재 연결, 유해 위험요인 인지 교육 및 안전 수칙
- 5번째 유사한 사고 (유사도: 0.9520)
- 사고원인: 작업자 위험작업 시 사전에 보고하지 아니하고, 생명줄 미설치 및 개인공구 점검불량으로 달비계와 메인
- ◆ 공종: 건축 > 지붕 및 홈통공사
- 작업프로세스: 설치작업
- ◆ 장소: 교육연구시설 / 외부
- ◆ 부위: 공구류 / 고소
- ➡ 대응 대책: 위험작업 시 사전알림 및 작업방법 안전교육, 개인보호구 및 개인공구 점검 철저, 건설장비 통행로 확보

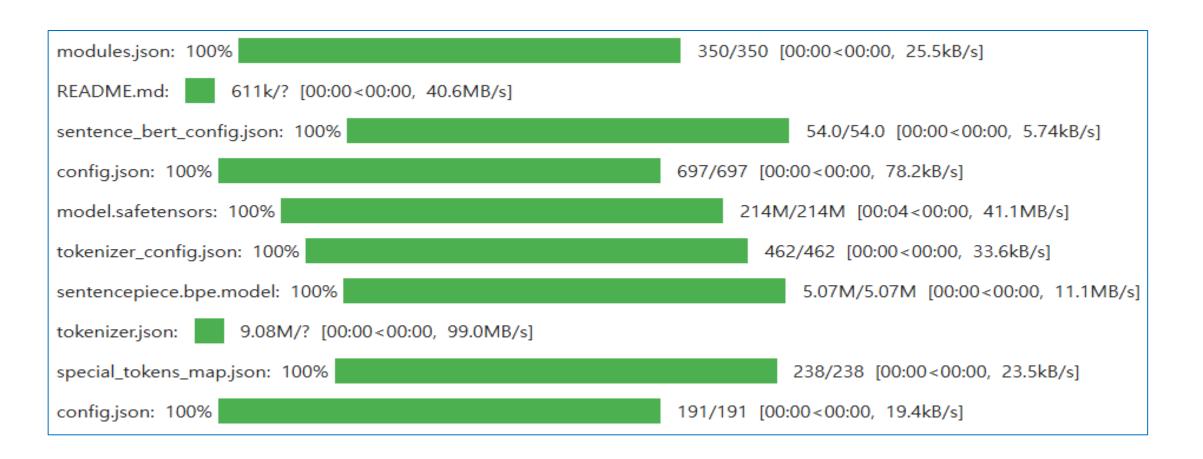
고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음.

기준 사고	안전장치 미흡	추락	
유사1	추락방호를 위한 작업 중	조치 미흡 및 관리감독 자 안전통제 미이행	0.9597
유사2	거푸집 해체공이 발판 해체 후	완전히 고정을 못함	0.9550
유사3	다른 자재를 가져오던 중 강 풍에 못이겨	판을 던지며 1층 바닥 으로 뛰어내림	0.9544
유사4	폼 판넬의 못을 제거하던 중	떨어지는 재해 발생	0.9527
유사5	생명줄 미설치 및 개인공구 점검불량으로	연결로프 노후로 끊어 찜	0.9520

### Part 4 HuggingFace: SentenceEmbedding



- Sentence Embedding
  - 기존 BERT보다 문장 간 비교에 최적화
  - BERT 기반으로 문장의 의미(semantic)를 벡터로 변환하는 딥러닝 모델
  - 다양한 사전학습 모델 제공 (all-MiniLM, paraphrase-MPNet, 등)
  - sentence-transformers 라이브러리에서 쉽게 사용 가능
  - model = SentenceEmbedding("ibm-granite/granite-embedding-107m-multilingual", device=device)





– 기존 BERT보다 문장 간 비교에 최적화, 임베딩만 필요 → HuggingFace 채택

항목	HuggingFace (SentenceTransformer)	LangChain
주요 목적	텍스트 벡터화	LLM 기반 파이프라인 내 임베딩/검색 등 통합 처리
추천 사용처	임베딩 자체가 중요한 경우	- RAG/Retrieval 기반 QA, 문서 검색+답변 통합 시스템
모델 종류	SBERT, KoBERT, GTE 등 수백 개 지원	기본적으로 OpenAI, HuggingFace, Cohere 임베딩을 래핑
장점	성능 우수, <b>다양한 한국어 모델 지원</b> , 로컬 추론	벡터저장소 통합, chain 구조 편리, RAG 시스템 구축에 적합
단점	벡터만 생성, 이후 검색/파이프라인은 직접 구현	모델 선택 유연성 낮음, 세부 제어 어려움
사용 방법	로컬 모델에 사용 가능	대부분 API 기반 (ex. OpenAl Embeddings)

# Part 4 HuggingFace: SentenceEmbedding

- ◆ 기준 사고원인: 고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음
- 기준 공종: 건축 > 철근콘크리트공사
- 기준 작업프로세스: 설치작업
- 기준 장소: 근린생활시설 / 내부
- 기준 부위: 철근 / 고소
- ➡ 대응 대책: 고소작업 시 추락 위험이 있는 부위에 안전장비 설치.
- ◆ 1번째 유사한 사고 (유사도: 0.9372)
- 사고원인: 타워크레인 8호기 상부 발판에서 해체작업을 위해 마스트 이동 중 미끄러지면서 좌측녹골을 부딪힘
- 공종: 기계설비 > 해체 및 철거공사
- 작업프로세스: 해체작업
- ◆ 장소: 공동주택 / 외부
- 부위: 타워크레인 / 상부(위)
- ➡ 대응 대책: 작업 전후 이동 시 안전사고 예방을 위한 안전교육 실시.
- 2번째 유사한 사고 (유사도: 0.9352)
- 사고원인: 거푸집 해제 작업 중 해제를 한번에 하려다 넘어진 사고
- 공종: 토목 > 항만공사
- ◆ 작업프로세스: 해체작업
- ◆ 장소: 계류시설 / 인접주변
- 부위: 거푸집 / 옆
- ➡ 대응 대책: 작업자 안전교육 및 작업 시 주의사항 전달을 통한 재발 방지 대책과 상태 지속확인 및 후속조치 검토클
- 3번째 유사한 사고 (유사도: 0.9345)
- 사고원인: 데크 철근사이에 발이 빠지면서 통증호소
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 타설작업
- ◆ 장소: 공장 / 내부
- ◆ 부위: 건물 / 바닥
- ➡ 대응 대책: 공사완료에 대한 재발 방지 대책과 향후 조치 계획
- 4번째 유사한 사고 (유사도: 0.9331)
- 사고원인: 거푸집 해체시 고정못 잡하다 못에 눈 맞음(보호안경 미착용, 작업미숙)
- 공종: 건축 > 철근콘크리트공사
- ◆ 작업프로세스: 해체작업
- ◆ 장소: 교육연구시설 / 내부
- 부위: 거푸집 / 상부(위)
- ➡ 대응 대책: 재발 방지 대책과 향후 조치 계획의 내용.
- 5번째 유사한 사고 (유사도: 0.9323)
- 사고원인: 폐목재 수거함에서 인양 작업 중에 크레인에 부착된 연결줄이 올라가는 과정에서 완전히 인양줄이 해체되
- ◆ 공종: 건축 > 철근콘크리트공사
- 작업프로세스: 정리작업
- ◆ 장소: 교육연구시설 / 외부
- 부위: 기타 / 바닥
- ➡ 대응 대책: 크레인 인양 작업 시 신호수 상시 배치와 안전사고 예방을 위한 사전교육 철저 및 현장 무인 크레인 사용

고소작업 중 추락 위험이 있음에도 불구하고, 안전난간대, 안전고리 착용 등 안전장치가 미흡하였음.

기준 사고	안전장치 미흡	추락	
유사1	8호기 상부 발판에서 해체 작업을 위해 마스트 이동 중	미끄러지면서 좌측늑 골을 부딪힘	0.9372
유사2	거푸집 해제 작업 중 해제를 한번에 하려다	넘어진 사고	0.9352
유사3	데크 철근사이에 발이 빠지 면서	통증호소	0.9345
유사4	거푸집 해체시 고정못 잡하 다	못에 눈 맞음	0.9331
유사5	완전히 인양줄이 해체되지 않은 상태에서 순간 충격으 로	작업자가 1.8미터 아 래로 뛰어 내림	0.9323



01

## 분석 데이터셋

- 공사현장 사고 데이터
- 자재관리 이력 데이터
- 근로자 출근 이력

02

전처리/가공

- 사전만들기
- 임베딩

02

사전학습된 KoBERT 임베딩 도입

02

사전 학습된 허깅페이스 임베딩 도입

**KoBERT** → **SentenceEmbedding** 

03-1

유사도 계산

- Cosine similarity
- TF-IDF

**Dot Product** 

03-2

가중치 계산

- MI 통계 기반 가중치
- 군집 기반 가중치

최종 유사도 행렬



## • 단순히 유사도가 높다고 좋은 모델이라고 할 수 있을까?

### ? 단순히 유사도가 높다고 좋은 모델일까?

아니요. 유사도 점수만으로 모델의 품질을 평가하는 것은 한계가 있습니다.

### ★ 왜 그런가?

- 1. 높은 유사도 = 의미 있는 관계?
- 코사인 유사도는 벡터 간 방향만 봅니다.
- 의미는 무시한 채 수치적으로만 가까운 경우도 있음.
- 예) "추락"과 "하늘" → 수치상 유사하나, 안전관리에서는 무관한 경우도 존재
- 2. 임베딩 자체의 왜곡 가능성
- 어떤 임베딩 모델은 비슷하지 않은 문장을 유사하게 출력할 수 있음.
- KoBERT처럼 L2 정규화된 모델은 유사도 값이 전체적으로 크게 나오는 경향도 있음.
- 3. 도메인 적합성 문제
- 일반 임베딩 모델은 건설사고 같은 특수 도메인에 약할 수 있음.
- 따라서 도메인 평가 지표(정확도, F1-score, 상관계수 등)가 함께 검토되어야 함.



점수 숫자만 보고 "어느 쪽이 더 좋다"고 단정할 수는 없다.

# 유사도 기준으로 산정한 가중치와 상관계수가 어느 정도 비례하는지 점검 필요

# Part 4 모델 검증 – KoBERT [2/3]

## • 컬럼 별 가중치 ↔ 최종 유사도행렬 상관계수

- 가중치 조합이 전체 유사도 행렬에 강하게 영향을 미치지 못하고 있음
- 가중치와 상관계수 순위는 대체로 일치하긴 하지만, 상관계수 절대값이 전반적으로 낮은 편임



컬럼	weights	상관계수	해석	검증
물적사고	0.2127	0.5372	가장 높은 가중치 → 상관계수도 1위	[일치]
인적사고	0.1353	0.4570	적당한 가중치, 적당한 영향력	[일치]
공종	0.1454	0.4535	예측 가능한 영향력	[일치]
사고객체	0.1078	0.4062	중간 수준	[대략일치]
장소	0.1182	0.3869	생각보다는 영향력이 낮음	[불일치]
부위	0.0814	0.3599	타당한 수준	[대략일치]
작업프로세스	0.1247	0.3587	낮은 편 (정보 중복 가능성?)	[불일치]
사고원인	0.0745	0.2084	낮은 가중치 ↔ 낮은 영향	[일치]

# Part 4 모델 검증 - SentenceEmbeding [3/3]



## • 컬럼 별 가중치 ↔ 최종 유사도행렬 상관계수

가중치를 줄 때 의도한 영향력과 실제 상관계수가 대체로 일치하거나 유사한 방향을 보이고 있음

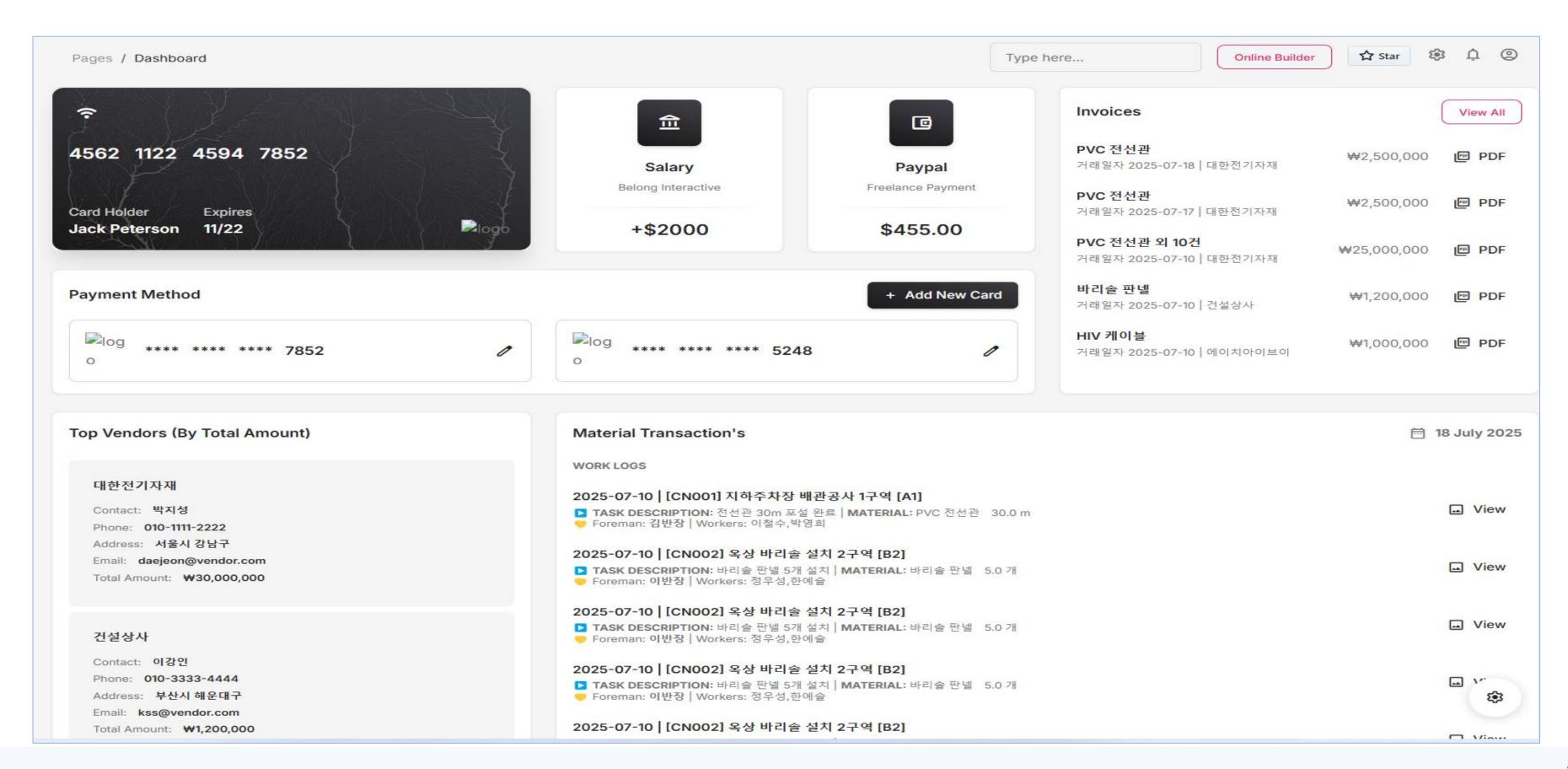


컬럼	weights	상관계수	해석	검증
사고객체	0.1078	0.8406	의외로 강한 영향 사고 분류에 있어 중요도가 높았을 가능성	[일치]
부위	0.0814	0.8316	비교적 낮은 가중치인데도 영향력 강함	[일치]
공종	0.1454	0.8224	상관 일치. 적절한 반영	[일치]
작업프로세스	0.1247	0.8171	일관된 영향력	[일치]
물적사고	0.2127	0.8158	가중치 1위였지만, 영향력은 약간 덜함	[대략일치]
인적사고	0.1353	0.8052	예상과 유사	[일치]
장소	0.1182	0.7694	중간 정도 영향력	[일치]
사고원인	0.0745	0.6425	가중치도 낮고, 실제로도 최하위	[일치]

# Part 05 서비스시연

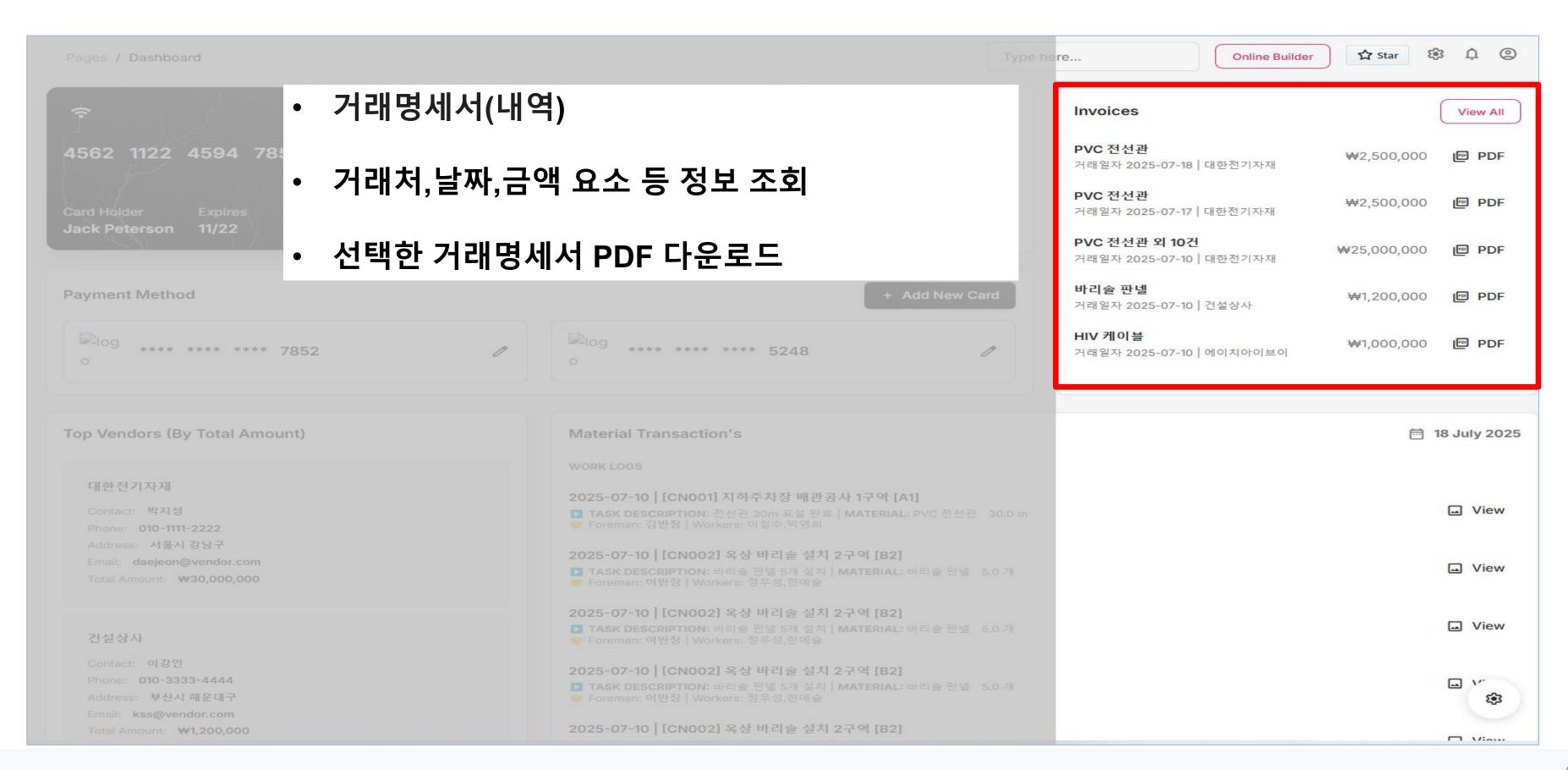
## Part 5 DashBoard





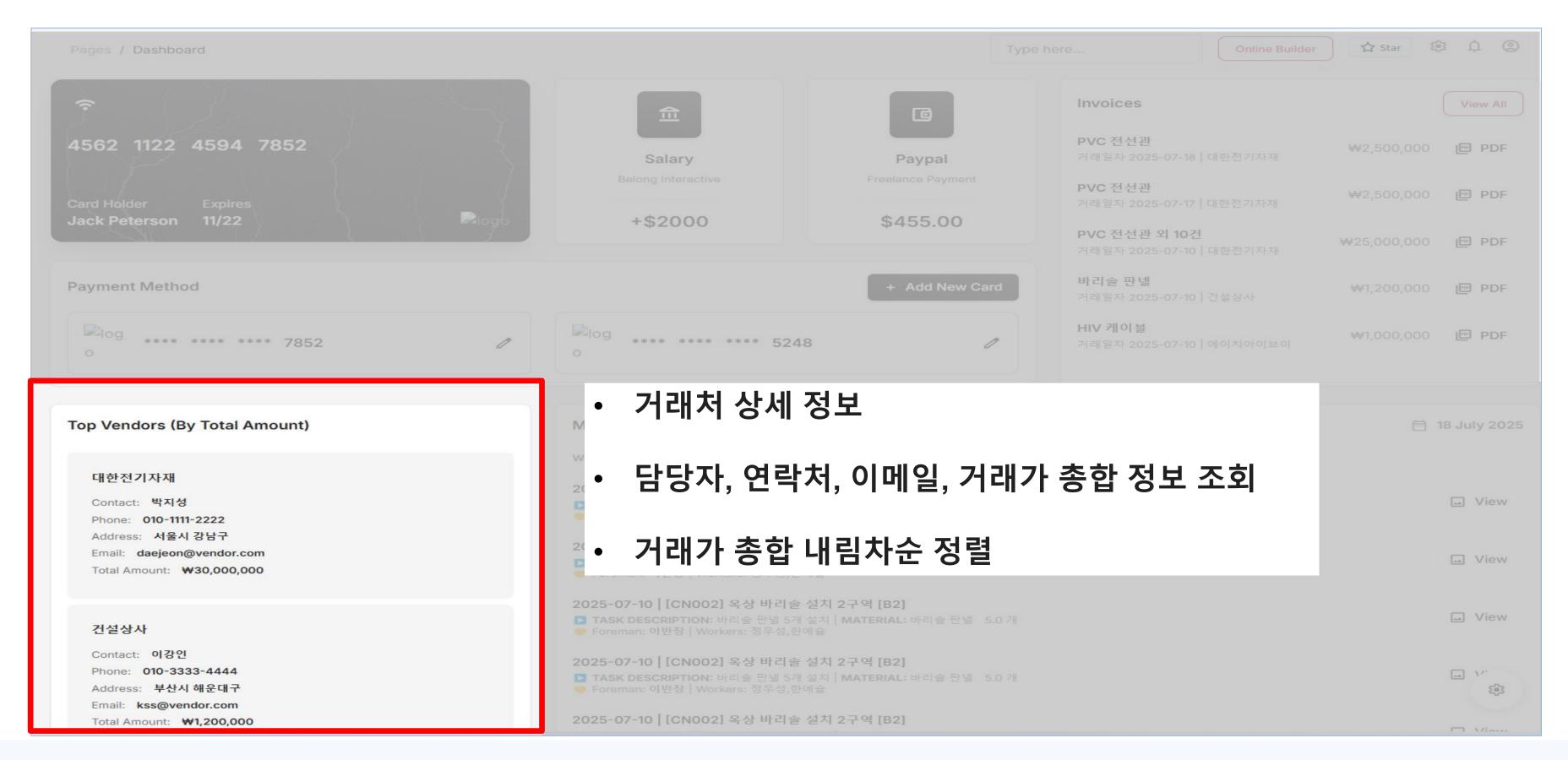
## Part 5 거래명세서





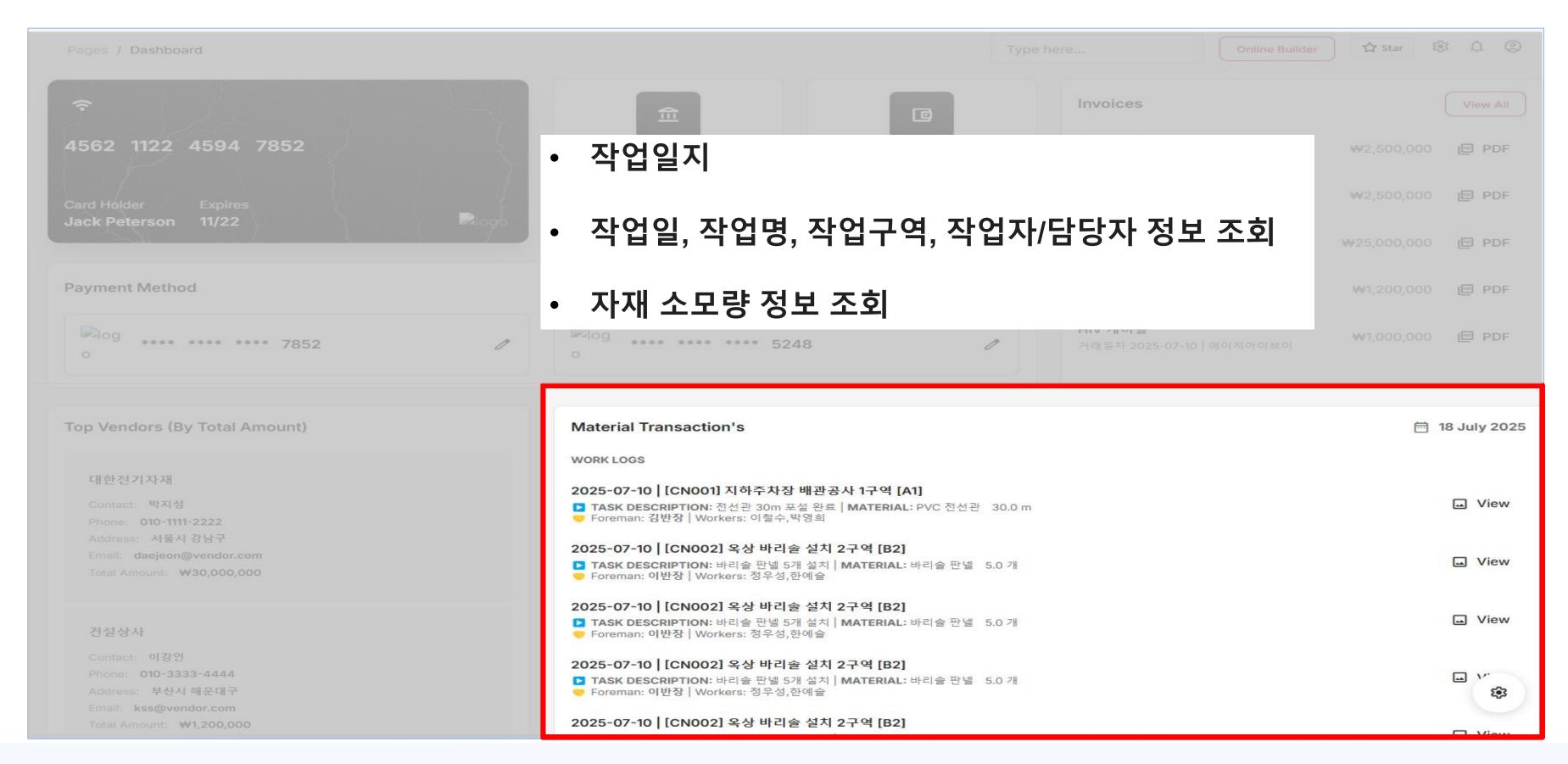
# Part 5 거래처 상세 정보





# Part 5 작업일지





# Part 5 사고 대응 매뉴얼 자동화 리포트



#### 사고 대응 매뉴얼 ■ 유사 사고 사례 1. 사고객체: 건설공구 > 공구류 사고원인: 보안경 미착용 및 불안전한 작업자세 공종: 기계설비 > 기계설비공사 작업프로세스: 조립작업 장소: 업무시설 / 내부 부위: 공구류 / 지하 대응 대책: 관련 근로자 보안경 일체형 안전모 지급 및 사용상태 확인을 통한 원인 규명과 재발 방지 대책 수립. 유사도: 0.8884 2. 사고객체: 건설공구 > 공구류 사고원인: 공도구사용 미숙에 따른 근로자의 불안전한 행동 공종: 건축 > 철근콘크리트공사 작업프로세스: 설치작업 장소: 교육연구시설 / 교육동 1층 거푸집작업 부위: 공구류 / 스킬 대응 대책: 공도구 작업 시 작업공간 확보, 사용 근로자에 대한 위험성 교육 및 공도구 점검 수시 확인, 슬라브 및 보 설치 작업 시 개인공도구에 대한 위험예 지활동 강화. 유사도: 0.8871 3. 사고객체: 건설기계 > 콘크리트펌프 사고원인: 작업자 개인 안전보호구 미착용 및 불안전한 행동 공종: 건축 > 철근콘크리트공사 작업프로세스: 타설작업 장소: 근린생활시설 / 내부 부위: 콘크리트펌프 / 옆 대응 대책: 현장책임자에 대한 관리 감독 철저와 유사사고 재발 방지를 위한 TBM시간 활용 안전교육 실시. 유사도: 0.8854 4. 사고객체: 건설공구 > 공구류 사고원인: 작업수칙 미준수 및 불안전한 행동 공종: 건축 > 목공사 작업프로세스: 절단작업 장소: 업무시설 / 내부 부위: 공구류 / 바닥 대응 대책: 사고 재발 방지 교육 실시와 위험성 평가 강화를 포함한 향후 조치 계획. 유사도: 0.8853 5. 사고객체: 건설자재 > 자재 사고원인: 안전수칙 미준수로 인한 작업자의 불안전한 행동 공종: 건축 > 해체 및 철거공사 작업프로세스: 이동 장소: 창고시설 / 내부 부위: 자재 / 옆 대응 대책: 안전수칙 준수에 따른 작업자의 안전 사항 철저함. 유사도: 0.8839

#### ■ 자동화 레포트

#### [분석]

기준 사고는 건설공구 중 공구류이며, 보안면 미착용과 숫돌 측면 사용이라는 인적요인이 핵심입니다. 유사사례들은 보호구 착용, 점검, 공간 확보, 교육을 강조합니다. 특히 사례 2번과 4번은 공간관리와 예지활동, 위험성 평가의 중요성을 보여줍니다.

#### [대응대책]

보안면 착용 의무화 및 사용 전 점검 시스템 도입, 핸드그라인더 작업 시 숫돌 정확한 위치 사용법 교육 강화, 현장 안전관리자 주관 수시 위험예지활동 실시 와 함께 공구 작업 공간 확보를 통한 재발 방지를 위한 종합안전점검 체계를 구축한다.

- 발생 사고 기준 과거 유사 사고 사례 Top-5 조회
- 사고 사례 토픽모델링
- 대책 및 대응 매뉴얼 리포트 생성 자동화
- 생성된 리포트 PDF 다운로드

# Part 5 근태관리



